

BBA 91626

Characterization of a complementary deoxyribonucleic acid for the coagulogen of *Limulus polyphemus*

Sheau-Mei Cheng, Akira Suzuki, Gerald Zon and Teh-Yung Liu

Division of Biochemistry and Biophysics, Office of Biologics Research and Review, Center for Drugs and Biologics,
Food and Drug Administration, 8800 Rockville Pike, Bethesda, MD 20892 (U.S.A.)

(Received 28 March 1986)

Key words: Coagulogen; Translation in vitro; Nucleotide sequence; RNA blot analysis; Southern blot analysis;
(*L. polyphemus*)

An 869-nucleotide-long cDNA clone for the coagulogen from *Limulus* amebocyte has been isolated and its nucleotide sequence has been determined. The deduced amino-acid sequence revealed a signal peptide, 20 amino acids long, and a mature protein of 175 amino acids. The amino-acid sequence of the coagulogen was compared to all known proteins by two computer programs. Using these programs, *Limulus* coagulogen showed 70% homology with the coagulogen of *Tachyplesus tridentatus* (Japanese horseshoe crab). Further computer analysis showed no statistically significant homology to support an evolutionary origin of the horseshoe crab coagulogen common to other protein families. These results place horseshoe crab coagulogen in a new superfamily unrelated to any other proteins investigated. RNA blot analysis of *Limulus* RNA indicated that the coagulogen mRNA was about 900 bases long and represented an abundant species in the amebocyte while detected only in small quantities in the hepatopancreas. Besides mature RNA, high-molecular-weight forms of coagulogen RNA were also observed. Southern blot analysis of *Limulus* DNA digested with restriction endonucleases suggested that the *Limulus* coagulogen gene contains at least three introns, or belongs to a multigene family.

Introduction

Limulus amebocyte lysate forms a gel when exposed to minute quantities of bacterial lipopolysaccharide [1]. This unique property has allowed the development of the lysate as an assay to detect contaminating pyrogens in pharmaceuticals and drugs intended for human use [2]. Our laboratory and others have been investigating the molecular

mechanisms involved in the *Limulus* clotting cascade system [3–13]. Endotoxin stimulates the amebocytes in the blood to undergo exocytosis, which releases the contents of their secretory granules to form a clot [10,11,13]. One of the secretory constituents of the granule is coagulogen, the clottable protein. This protein has been purified [5] and its primary structure was recently elucidated [14]. The nature of the mRNA that codes for the synthesis of coagulogen has not been studied.

In this report, we describe the use of recombinant DNA technology to isolate and characterize the cDNA of the *Limulus polyphemus* coagulogen. These results further facilitated the study of the coagulogen at both the protein and gene level.

Correspondence address: Dr. T.-Y. Liu, Division of Biochemistry and Biophysics, Office of Biologics Research and Review, Center for Drugs and Biologics, Food and Drug Administration, 8800 Rockville Pike, Bethesda, MD 20892, U.S.A.

Materials and Methods

Materials. *Escherichia coli* DNA polymerase I (Klenow fragment), nuclease S_1 , proteinase K, terminal deoxynucleotidyl transferase and T_4 -DNA ligase were purchased from Boehringer-Mannheim Biochemicals. Restriction endonucleases, T_4 -polynucleotide kinase, M13 vectors (mp18 and mp19), Immunoprecipitin, nick-translation system and the dG-tailed pBR322 vector were obtained from Bethesda Research Laboratory. Rabbit reticulocyte lysate and RNasin were from Promega Biotech. RNAase A and dextran sulfate were purchased from Sigma Chemical Corp. Guanidine thiocyanate was from Fluka, Switzerland. Nonradioactive nucleotide triphosphates were purchased from P-L Biochemicals. [α - ^{32} P]dATP about 400 Ci/mmol, [α - ^{32} P]dCTP, [α - ^{32} P]dTTP, and [γ - ^{32} P]ATP at about 3000 Ci/mmol were purchased from New England Nuclear. [^{35} S]Methionine at about 1000 Ci/mmol was from Amersham.

In vitro translation of *Limulus* coagulogen. *Limulus* embocyte poly (A)⁺ RNA (0.5 μ g) was translated in a rabbit reticulocyte lysate containing [^{35} S]methionine (about 1000 Ci/mmol). The translation products were immunoprecipitated using a coagulogen-specific antiserum [5] * and Immunoprecipitin. The size of radiolabeled coagulogen was estimated using 13% sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis according to Laemmli [15].

Nucleic acid preparation. *Limulus* amebocyte cells were collected from horseshoe crabs by cardiac puncture [16]. The blood, collected into ice-cold containers, was centrifuged at 1000 rpm for 15 min at 4°C. The pellet was washed three times with ice-cold 3% saline and lysed with 4 M guanidium thiocyanate [17]. The poly(A)⁺ RNA fractions were prepared using an oligo(dT)-cellulose column [18]. The high-molecular-weight DNA was prepared as follows. The washed cell pellet obtained from 30 ml of *Limulus* amebocyte was gently resuspended in a 10 ml lysis solution

(500 mM EDTA (pH 8.0), 0.5% sarkosyl and 100 μ g/ml proteinase K) and incubated at 50°C for 3 h. The resulting lysate was extracted three times with phenol and dialyzed against 10 mM Tris (pH 8.0)/1 mM EDTA (buffer 1) overnight. In addition, the lysate was treated with RNAase A (50 μ g/ml) and proteinase K (100 μ g/ml) at 37°C for 3 h, extracted three times with phenol, and dialyzed against buffer 1 overnight.

***Limulus* cDNA library.** An *L. polyphemus* cDNA library was constructed as previously described [19], except for the following modifications. In the presence of ribonuclease inhibitor (RNasin, 1650 units/ml), synthesis of the first cDNA strand was accomplished by using *Limulus* poly (A)⁺ RNA as the template and the avian myeloblastosis virus (AMV) reverse transcriptase (100 units/ μ g of template, at pH 8.3). The reaction mixture was incubated at 37°C for 1 h. The newly synthesized single-stranded cDNA was separated from mRNA (boiled for 3 min and chilled on ice for 10 min) and used to prime the synthesis of the second cDNA strand in the presence of the Klenow fragment of *E. coli* DNA polymerase I. The resulting double-stranded cDNA was cleaved with nuclease S_1 , and then fractionated into size classes by passage through a Sepharose CL-4B column. Size-selected molecules of double-stranded DNA greater than approx. 500 base pairs were tailed with poly(dC) in the presence of terminal deoxynucleotidyl transferase and inserted into the pBR322 vector that had been cleaved with *Pst*I and tailed with dG. The chimeric plasmids were then introduced into *E. coli* strain MC1061 [20] to yield the *L. polyphemus* cDNA library.

Preparation of oligonucleotide probes. Two single-stranded DNA probes (Fig. 1) specific for *Limulus* coagulogen were prepared by the phosphoramidite synthesis procedure using an Applied Biosystems DNA synthesizer (Foster City, CA) [21,22,23]. The sequences of these probes were based on the primary amino-acid sequences of a stretch of 10 amino acids within a 27-amino-acid peptide (peptide C, amino-acid residues 19–46) released from the coagulogen upon gelation [14]. Probes were labeled at their 5' end using T_4 polynucleotide kinase, [γ - ^{32} P]ATP and the procedure of Maxam and Gilbert [24] to a specific activity of approx. $(1-2) \cdot 10^6$ dpm/pmol.

* Unpublished work (Liu, T.-Y.). Antibodies against coagulogen were prepared in rabbit using purified coagulogen [5].

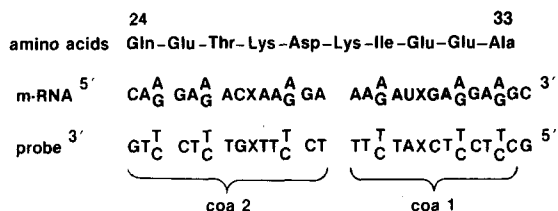


Fig. 1. Mixed oligonucleotide probes. The amino-acid sequences from *Limulus* coagulogen [36] used to specify the two 14-mer oligonucleotides (residues 24–28, residues 29–33), the possible codons for those amino acids and the sequences of the oligonucleotide probes designated to account for all possible combinations are shown.

Screening recombinant colonies by hybridization. The colony hybridization procedure of Grunstein and Hogness was used [25]. 2000 tetracycline-resistant colonies were screened for the *Limulus* coagulogen-specific cDNA inserts using the two 5'-end ³²P-labeled synthetic oligonucleotides as probes (Fig. 1). Hybridizations were carried out at 35°C for 16 h in a solution containing 5 × SSC (1 × SSC = 0.15 M sodium chloride/0.015 M sodium citrate), 5 × Denhardt's solution (1 × Denhardt's solution = 0.02% Ficoll, 0.02% poly(vinylpyrrolidone), 0.02% bovine serum albumin [26]), 0.1% SDS, 10% (w/v) dextran sulfate, 100 µg/ml sonicated, denatured salmon-sperm DNA, and (1–2) · 10⁷ dpm ³²P-labeled probe. Filters were washed four times with 1 × SSC plus 0.1% SDS at room temperature for 1 h, blotted dry and autoradiographed at –70°C with intensifying screens.

Sequencing of cDNA clones. The cDNA inserts

of two selected plasmids (pcoa 1 and pcoa 5) were ligated to the replicative form of M13 vectors (mp18 and mp19) after cleavage with the restriction endonuclease *Pst*I. Competent *E. coli* cells (strain JM103) were transfected with the recombinant M13 vectors. Single-stranded cDNA from individual plaques were prepared and used as templates for the chain termination sequencing procedure of Sanger [27]. In Fig. 2, the strategies used for the sequencing of both strands of the cDNA inserts are shown.

RNA blot analysis. RNA samples were denatured and fractionated on a 1.3% agarose gel containing formaldehyde as described by Goldberg [28]. The RNA was transferred to nitrocellulose paper and hybridized with nick-translated pcoa 1 as described by Wahl et al. [29]. Following hybridization the filter was washed in 0.1 × SSC/0.1% SDS at room temperature and autoradiographed using Kodak XAR-5 film at –70°C. The ribosomal RNA of *E. coli* (23 S, 16 S) and mouse (28 S and 18 S), and rabbit globin mRNA (9 S) were used as unlabeled RNA size markers on the same gel [30,31].

Southern blot analysis. DNA was digested to the completion with restriction enzymes (*Eco*RI, *Eco*RI + *Bam*HI and *Hind*III), electrophoresed through a 0.9% agarose gel, transferred to a nitrocellulose filter [32], and hybridized with the nick-translated ³²P-labeled cDNA insert of pcoa 1. Following the hybridization, the filter was washed as described previously [33].

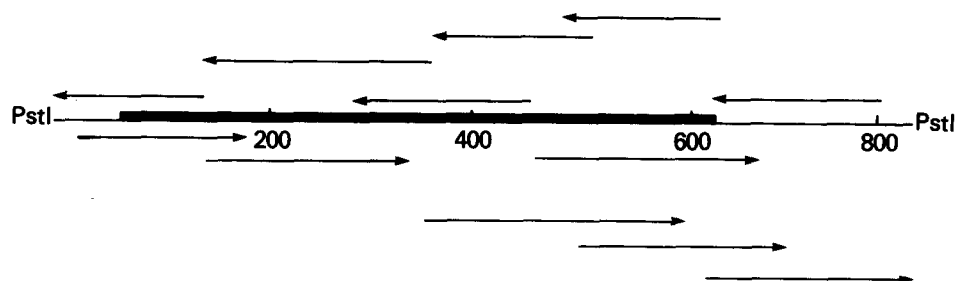


Fig. 2. Strategy used to determine the nucleotide sequence of the cDNA clones. The direction of mRNA transcription is from left to right. Initially, sequencing was performed using the universal M13 15-mer primer. Subsequently, new primers were synthesized according to the previously identified DNA sequence. The back end of the arrow indicates the position of the primer used in the dideoxy method of DNA sequencing. The arrows indicate the direction of sequence analysis (5' → 3') and the length of sequence obtained. Thick line represents the protein coding region.

Results

Cell free synthesis of pre-coagulogen

To study the newly synthesized coagulogen, a rabbit reticulocyte lysate system was used in the presence of *Limulus* mRNA. The presence of a precursor form of *Limulus* coagulogen in the rabbit reticulocyte lysate system is shown in lane 1, Fig. 3. Immunoprecipitation of this precursor-form of coagulogen could be effectively inhibited by an excess (20 μ g) of unlabeled purified coagulogen (lane 2, Fig. 3), confirming that the radiolabeled protein band shown in lane 1, Fig. 3 is related to the *Limulus* coagulogen. On the basis of relative mobility, it is estimated that the precursor-form of the *Limulus* coagulogen is approx. 2000 Da larger than the mature protein (lane 3, Fig. 3).

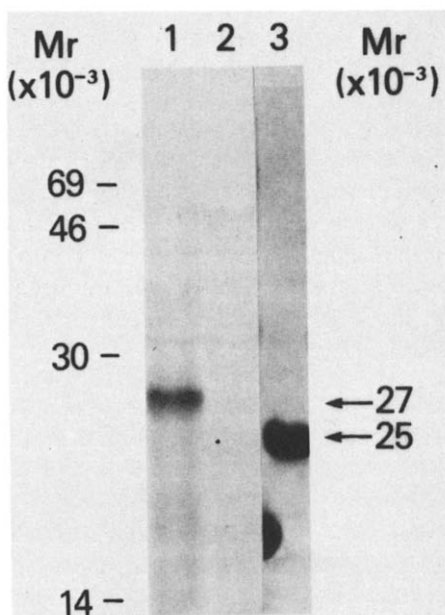


Fig. 3. Cell-free translation of *Limulus* mRNA products. Translation product after precipitation with antibody and Immunoprecipitin is shown in lane 1. Immunoprecipitation of total translation product mixed with unlabeled authentic coagulogen (20 μ g) is shown in lane 2. The purified mature protein is shown in lane 3. The size of coagulogen was estimated using 13% SDS polyacrylamide gel electrophoresis according to Laemmli [15]. The size markers are shown on the left side of the figure.

Isolation and characterization of *Limulus* coagulogen cDNA clones

The *Limulus* cDNA library was screened for the presence of the coagulogen coding sequences. Of the approx. 2000 tetracycline-resistant colonies screened, 8 colonies hybridized with the two ^{32} P-labeled probes each 32-fold degenerate in sequence. Results of restriction enzyme mapping indicated that the cDNA inserts of these 8 clones shared a resemblance (data not shown). The clone containing the longest insert, designated as pcoa 1, was 840 bp long. One of the remaining clones, designated as pcoa 5, was 555 bp in size. To obtain the nucleotide sequence of these two cDNA clones, both strands of the individual clones were sequenced by the dideoxy method [27]. The entire cDNA sequence and the predicted amino-acid sequence of coagulogen are shown in Fig. 4. The combined nucleotide sequence derived from both pcoa 1 and pcoa 5 was 869 bp in length, including 110 bp of 5' noncoding region, 525 bp in the coding region, and 234 bp in the 3' noncoding region following the TAA termination codon. Neither the poly(A) tail nor the AATAAA sequence was included in these cDNA clones. The first amino acid of the mature protein is a glycine [3,14]. The sequence preceding this amino acid appears to be a 20-amino-acid hydrophobic leader peptide beginning with methionine. The overlapping nucleotide sequences between cDNA inserts of pcoa 1 and pcoa 5 were completely identical. The beginning and the end positions of the insert of plasmid pcoa 5 are shown in Fig. 4. This plasmid begins 29 bp further upstream than pcoa 1. Two termination codons were observed in these 29 nucleotides. This provides additional support for the contention that the first methionine of the leader peptide is the initiation codon for amino-acid translation.

Analysis of *Limulus* RNA

To examine the concentration of coagulogen specific mRNA in the *Limulus* amoebocyte and hepatopancreas, radiolabeled *Limulus* coagulogen-specific cDNA insert from pcoa 1 was used as a probe. The results of the RNA blot analysis are shown in Fig. 5. One prominent band of approx. 900 bases in size was found to be abundant in samples of total cellular RNA (lane 1, Fig. 5) and

CT	CCA	TTC	TGA	CGT	CAG	GCT	20
*							
-30	TGT	TAG	AGA	AAG	AAG	AAA	GCT
-20	Met	Glu	Lys	Lys	Lys	AAA	TTG
-10	Leu	Leu	Leu	Leu	Leu	TTA	TTG
1	Gly	Asp	Pro	Asn	Val	GTC	CCA
10	Gly	Asp	Pro	Asn	Val	GTC	CCA
20	Gly	Asp	Pro	Asn	Val	GTC	CCA
30	Gly	Asp	Pro	Asn	Val	GTC	CCA
40	Gly	Asp	Pro	Asn	Val	GTC	CCA
50	Gly	Asp	Pro	Asn	Val	GTC	CCA
60	Gly	Asp	Pro	Asn	Val	GTC	CCA
70	Gly	Asp	Pro	Asn	Val	GTC	CCA
80	Gly	Asp	Pro	Asn	Val	GTC	CCA
90	Gly	Asp	Pro	Asn	Val	GTC	CCA
100	Gly	Asp	Pro	Asn	Val	GTC	CCA
110	Gly	Asp	Pro	Asn	Val	GTC	CCA
120	Gly	Asp	Pro	Asn	Val	GTC	CCA
130	Gly	Asp	Pro	Asn	Val	GTC	CCA
140	Gly	Asp	Pro	Asn	Val	GTC	CCA
150	Gly	Asp	Pro	Asn	Val	GTC	CCA
160	Gly	Asp	Pro	Asn	Val	GTC	CCA
170	Gly	Asp	Pro	Asn	Val	GTC	CCA
180	Gly	Asp	Pro	Asn	Val	GTC	CCA
190	Gly	Asp	Pro	Asn	Val	GTC	CCA
200	Gly	Asp	Pro	Asn	Val	GTC	CCA
210	Gly	Asp	Pro	Asn	Val	GTC	CCA
220	Gly	Asp	Pro	Asn	Val	GTC	CCA
230	Gly	Asp	Pro	Asn	Val	GTC	CCA
240	Gly	Asp	Pro	Asn	Val	GTC	CCA
250	Gly	Asp	Pro	Asn	Val	GTC	CCA
260	Gly	Asp	Pro	Asn	Val	GTC	CCA
270	Gly	Asp	Pro	Asn	Val	GTC	CCA
280	Gly	Asp	Pro	Asn	Val	GTC	CCA
290	Gly	Asp	Pro	Asn	Val	GTC	CCA
300	Gly	Asp	Pro	Asn	Val	GTC	CCA
310	Gly	Asp	Pro	Asn	Val	GTC	CCA
320	Gly	Asp	Pro	Asn	Val	GTC	CCA
330	Gly	Asp	Pro	Asn	Val	GTC	CCA
340	Gly	Asp	Pro	Asn	Val	GTC	CCA
350	Gly	Asp	Pro	Asn	Val	GTC	CCA
360	Gly	Asp	Pro	Asn	Val	GTC	CCA
370	Gly	Asp	Pro	Asn	Val	GTC	CCA
380	Gly	Asp	Pro	Asn	Val	GTC	CCA
390	Gly	Asp	Pro	Asn	Val	GTC	CCA
400	Gly	Asp	Pro	Asn	Val	GTC	CCA
410	Gly	Asp	Pro	Asn	Val	GTC	CCA
420	Gly	Asp	Pro	Asn	Val	GTC	CCA
430	Gly	Asp	Pro	Asn	Val	GTC	CCA
440	Gly	Asp	Pro	Asn	Val	GTC	CCA
450	Gly	Asp	Pro	Asn	Val	GTC	CCA
460	Gly	Asp	Pro	Asn	Val	GTC	CCA
470	Gly	Asp	Pro	Asn	Val	GTC	CCA
480	Gly	Asp	Pro	Asn	Val	GTC	CCA
490	Gly	Asp	Pro	Asn	Val	GTC	CCA
500	Gly	Asp	Pro	Asn	Val	GTC	CCA
510	Gly	Asp	Pro	Asn	Val	GTC	CCA
520	Gly	Asp	Pro	Asn	Val	GTC	CCA
530	Gly	Asp	Pro	Asn	Val	GTC	CCA
540	Gly	Asp	Pro	Asn	Val	GTC	CCA
550	Gly	Asp	Pro	Asn	Val	GTC	CCA
560	Gly	Asp	Pro	Asn	Val	GTC	CCA
570	Gly	Asp	Pro	Asn	Val	GTC	CCA
580	Gly	Asp	Pro	Asn	Val	GTC	CCA
590	Gly	Asp	Pro	Asn	Val	GTC	CCA
600	Gly	Asp	Pro	Asn	Val	GTC	CCA
610	Gly	Asp	Pro	Asn	Val	GTC	CCA
620	Gly	Asp	Pro	Asn	Val	GTC	CCA
630	Gly	Asp	Pro	Asn	Val	GTC	CCA
640	Gly	Asp	Pro	Asn	Val	GTC	CCA
650	Gly	Asp	Pro	Asn	Val	GTC	CCA
660	Gly	Asp	Pro	Asn	Val	GTC	CCA
670	Gly	Asp	Pro	Asn	Val	GTC	CCA
680	Gly	Asp	Pro	Asn	Val	GTC	CCA
690	Gly	Asp	Pro	Asn	Val	GTC	CCA
700	Gly	Asp	Pro	Asn	Val	GTC	CCA
710	Gly	Asp	Pro	Asn	Val	GTC	CCA
720	Gly	Asp	Pro	Asn	Val	GTC	CCA
730	Gly	Asp	Pro	Asn	Val	GTC	CCA
740	Gly	Asp	Pro	Asn	Val	GTC	CCA
750	Gly	Asp	Pro	Asn	Val	GTC	CCA
760	Gly	Asp	Pro	Asn	Val	GTC	CCA
770	Gly	Asp	Pro	Asn	Val	GTC	CCA
780	Gly	Asp	Pro	Asn	Val	GTC	CCA
790	Gly	Asp	Pro	Asn	Val	GTC	CCA
800	Gly	Asp	Pro	Asn	Val	GTC	CCA
810	Gly	Asp	Pro	Asn	Val	GTC	CCA
820	Gly	Asp	Pro	Asn	Val	GTC	CCA
830	Gly	Asp	Pro	Asn	Val	GTC	CCA
840	Gly	Asp	Pro	Asn	Val	GTC	CCA
850	Gly	Asp	Pro	Asn	Val	GTC	CCA
860	Gly	Asp	Pro	Asn	Val	GTC	CCA
870	Gly	Asp	Pro	Asn	Val	GTC	CCA
880	Gly	Asp	Pro	Asn	Val	GTC	CCA
890	Gly	Asp	Pro	Asn	Val	GTC	CCA
900	Gly	Asp	Pro	Asn	Val	GTC	CCA
910	Gly	Asp	Pro	Asn	Val	GTC	CCA
920	Gly	Asp	Pro	Asn	Val	GTC	CCA
930	Gly	Asp	Pro	Asn	Val	GTC	CCA
940	Gly	Asp	Pro	Asn	Val	GTC	CCA
950	Gly	Asp	Pro	Asn	Val	GTC	CCA
960	Gly	Asp	Pro	Asn	Val	GTC	CCA
970	Gly	Asp	Pro	Asn	Val	GTC	CCA
980	Gly	Asp	Pro	Asn	Val	GTC	CCA
990	Gly	Asp	Pro	Asn	Val	GTC	CCA
1000	Gly	Asp	Pro	Asn	Val	GTC	CCA

Fig. 4. The nucleotide sequence of *Limulus* coagulogen cDNA clones. The nucleotide sequence of *Limulus* coagulogen cDNA and the deduced amino-acid sequence are shown. Amino acids are numbered according to the *Limulus* coagulogen protein sequence of Miyata et al., [14] and shown above the coding portion of the cDNA. Nucleotides are numbered on the right of the figure. Residues determined by Miyata et al., that differ from the deduced amino-acid sequence are indicated with parentheses. The limits of pcoa 1 sequence are marked with arrows, and the limits of pcoa 5 sequence are marked with asterisks. The sequence underlined by a solid line in the coding region corresponds to the two oligonucleotide probes (Fig. 1) used for screening the cDNA library. Boxes indicate in frame termination codons mentioned in the text.

in poly(A)⁺ fraction (lane 3, Fig. 5) of the *Limulus* amoebocyte. However, this species was detected only in low quantity in samples of total cellular RNA from *Limulus* hepatopancreas (lane 2, Fig. 5). These results indicate that the sequence given in Fig. 4 represents a nearly complete cDNA sequence missing a total of approx. 30 nucleotides at the 3' and 5' ends.

In addition to the major species of coagulogen-specific mRNA identified, at least three discrete higher-molecular-weight RNAs that hybridized with the radiolabeled clone pcoa 1 were noted in the amoebocyte total RNA (lane 1, Fig. 5). These may represent either cross-hybridization of the

coagulogen probe with an RNA species encoding other proteins or more likely precursor forms of the major coagulogen-specific mRNA.

Southern analysis of *Limulus* genomic DNA

To determine the structural organization of the *Limulus* coagulogen gene, total genomic DNA was digested to completion with various restriction endonucleases, and hybridized with the ³²P-labeled cDNA probe. The hybridization patterns of Southern analysis show at least two *Eco*RI fragments (Lane 1, Fig. 6) and four *Hind*III fragments (Lane 3, Fig. 5). Because the DNA sequence of the cDNA pcoa 1 contains neither a *Hind*III nor a *Eco*RI site, a single hybridization fragment should be obtained in both the *Hind*III and *Eco*RI digests of total genomic DNA if coagulogen was coded by a single gene without introns. Therefore, our results suggested that the *Limulus* coagulogen gene contains at least three introns, or belongs to a multigene family.

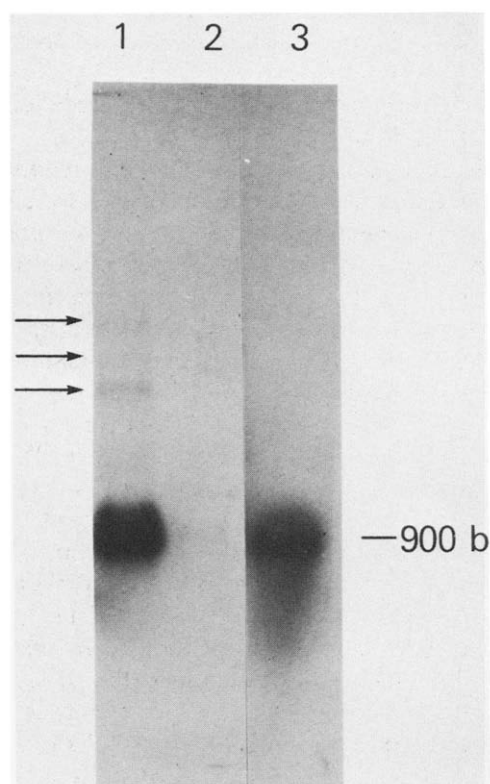


Fig. 5. RNA blot analysis of *Limulus* coagulogen-specific RNA. Lane 1 contains *Limulus* amoebocyte total cellular RNA (10 µg), lane 2 contains *Limulus* hepatic total cellular RNA (10 µg), and lane 3 contains *Limulus* amoebocyte poly(A)⁺ RNA (1 µg). A 1.3% agarose-formaldehyde gel electrophoresis was performed. The RNA blots were hybridized with nick-translated pcoa 1. Three arrows emphasize the weak hybridization signals. The size marker is shown on the right-hand side of the figure.

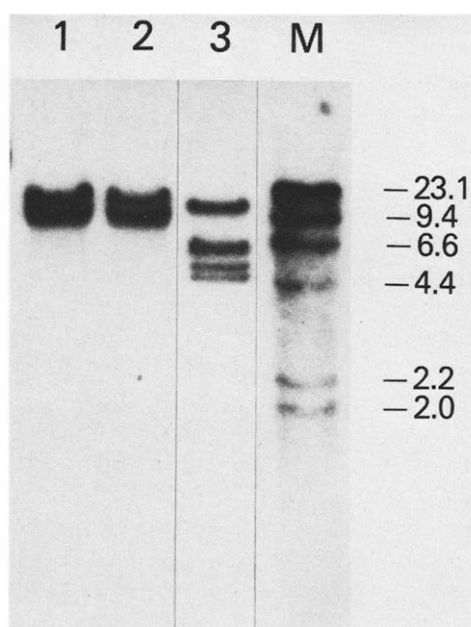


Fig. 6. Detection of *Limulus* coagulogen gene in the total genomic DNA. 10 µg of *Limulus* genomic DNA was digested to completion with either *Eco*RI (lane 1), *Eco*RI + *Bam*HI (lane 2) or *Hind*III (lane 3) and resolved on a 0.9% agarose gel. The nick-translated cDNA insert of pcoa 1 was used as the hybridization probe. The size markers are λ -*Hind*III digests. The numbers listed on the right-hand side are the positions of size markers.

Discussion

The nucleotide sequence for the coding region as well as most of the 5' and 3' untranslated region of *Limulus* coagulogen cDNA has been determined. This study represented the first isolation and characterization of a cDNA clone from the ancient arachnoid *L. polyphemus*. The results showed that the codon usage pattern for the *Limulus* coagulogen is biased and is similar to those of other eucaryotic proteins [34]. Codons containing the dinucleotide CG (e.g., UCG for serine, ACG for threonine, GCG for alanine, CCG for proline and CGN for arginine) are least utilized.

The amino-acid coding region consists of 525 nucleotides covering the full length of the mature protein of 175 amino acids plus an additional 60 nucleotides coding for 20 amino acids of a presumptive leader sequence. The existence of a pre-coagulogen with a molecular weight of about 200 Da (approx. 20 amino acids) larger than the mature protein was predicted from the in vitro translation data (Fig. 3). The amino-acid sequence of the presumptive leader peptide possesses the general properties described for signal peptides [35] including the existence of a long stretch of hydrophobic amino acids located between a cluster of basic amino acids (Lys-Lys) and the cleavage site (alanine at position -1). Previous studies [11-13] have shown that mature coagulogen molecules are stored in the secretory granules of the resting *Limulus* amoebocytes. Presumably, pre-coagulogen is processed as it passes through the lumen of endoplasmic reticulum and Golgi apparatus, and the mature protein is then packaged into the secretory granules. Upon exposure of the amoebocyte to endotoxin, the granules fuse with cell membranes resulting in the expulsion of the constituents of the granules including the coagulogen. In addition to coagulogen, *Limulus* amoebocyte secretory granules contain the proclotting enzyme [11-13] and a number of other proteins [11-13] associated with the clotting cascade system.

Among the constituents of the secretory granules in the *Limulus* amoebocyte, the coagulogen is the major soluble protein. It accounts for 40% of the total soluble lysate protein of the amoebocyte [36]. The results of in vitro translation studies in

the presence of [³⁵S]methionine indicated that the pre-coagulogen represents approx. 0.5% of the total translatable proteins. These data are in accordance with the finding that 8 out of 2000 (0.4%) recombinant cDNA clones possessed the coagulogen specific sequences. RNA blot analysis of the *Limulus* RNA showed that coagulogen is expressed predominantly in the amoebocyte rather than the hepatopancreas. Southern blot analysis of the amoebocyte DNAs indicated that the coagulogen contained at least three introns, or belongs to a multigene family.

The predicted amino-acid sequence of coagulogen, derived from the cDNA sequence reported in this study, is in good agreement with the results obtained by proteolytic digestion of the coagulogen [14], while our data were derived from analysis of both strands of DNA obtained from two independent cDNA clones. Therefore, we are proposing that the amino-acid sequence predicted from DNA analysis would be the correct one.

The deduced amino-acid sequence of mature *Limulus* coagulogen has been compared to all known proteins whose primary structures were published before August 1985. The comparison was carried out by two computer programs, SEARCH and RELATE, developed by Dayhoff and Barker [40]. In the first instance, 27-amino-acid sequence segments of *Limulus* coagulogen were compared to all 2784 protein sequences in the data collection. In the second instance, the entire sequence of *Limulus* coagulogen was compared against all other proteins whose sequences are known. Except for the coagulogen of *Tachyplesus* (a Japanese horseshoe crab) [14] these comparisons did not reveal any statistically meaningful homology with any other proteins, including the primate fibrinopeptides. Previously, it was pointed out that homology existed between residues 1-18 and residues 39-46 of *Limulus* coagulogen with primate fibrinopeptide B [14]. However, our computer analysis indicates that the 'limited homology' noted cannot be regarded as sufficient evidence to support a common evolutionary origin of coagulogen and the primate fibrinopeptides. At this point in time, we believe it is preferable to assign *Limulus* and *Tachyplesus* coagulogen to a different superfamily unrelated to any other proteins.

In summary, the isolation and characterization of the *Limulus* coagulogen cDNA make it possible to elucidate the genomic organization of coagulogen gene. The isolation of this gene from a *Limulus* genomic library is in progress. These studies provide the initial step to examine the clottable components involved in the *Limulus* coagulation process at gene level. Hopefully this will lead to the further understanding of the development of clotting systems.

Acknowledgements

We wish to thank Dr. W. Barker for performing computer analysis, Kathy Gallo and Carrie Samson for providing us with oligodeoxyribonucleotide probes, and Kevin Nahigian for technical assistance during the course of this work.

References

- * 1 Levin, J. and Bang, F.B. (1964) Bull. Johns Hopkins Hosp. 115, 265-274
- 2 Cooper, J.F., Hochstein, H.D. and Seligmann, E.B., Jr. (1972) Bulletin Parenter, Drug Assoc. 26, 153-162
- 3 Tai, J.Y. and Liu, T.-Y. (1977) J. Biol. Chem. 252, 2178-2181
- 4 Liang, S.M., Lamar, T.P. and Liu, T.-Y. (1980) J. Biol. Chem. 255, 5586-5590
- 5 Tai, J.Y., Seid, R.C., Jr., Huhn, R.D. Liu, T.-Y. (1977) J. Biol. Chem. 252, 4773-4776
- 6 Ohki, M., Nakamura, T., Morita, T. and Iwanga, S. (1980) FEBS Lett. 120, 217-220
- 7 Solum, N.O. (1973) Thromb. Res. 2, 55-70
- 8 Nakamura, S. and Levin, J. (1982) Biochim. Biophys. Acta 707, 217-225
- 9 Nakamura, S. and Levin, J. (1982) Biochem. Biophys. Res. Commun. 108, 1619-1623
- 10 Dumont, J.N., Anderson, E. and Winner, G. (1966) J. Morphol. 119, 181-208
- 11 Ornberg, R.L. and Reese, T.S. (1981) J. Cell Biol. 90, 40-54
- 12 Mürer, E.H., Levin, J. and Home, R. (1975) J. Cell Physiol. 86, 535-542
- 13 Liang, S.M., Hascall, G., Liu, T. and Liu, T.-Y. (1985) J. Protein Chem. 4, 151-162
- 14 Miyata, T., Hiranga, M., Umezu, M. and Iwanaga, S. (1984) J. Biol. Chem. 259, 8924-2833
- 15 Laemmli, U.K. (1970) Nature 222, 680-685
- 16 Jorgensen, J.H. and Smith, R.F. (1973) Appl. Microbiol. 26, 43-48
- 17 Glisin, V., Crkvenjakov, R. and Byus, C. (1974) Biochemistry 18, 5294-5299
- 18 Aviv, H. and Leder, P. (1972) Proc. Natl. Acad. Sci. USA 69, 1408-1412
- 19 Goodman, H.M. and MacDonald, R.J. (1979) Methods Enzymol. 68, 75-83
- 20 Boyer, H.W. and Roulland-Dussoix (1969) J. Mol. Biol. 41, 459-465
- 21 Caruthers, M.H., Beaucage, S.L., Becker, C., Efcavitch, W., Fisher, E.F., Galluppi, G., Goldman, R., De Haseth, P., Martin, F., Mattuecci, M. and Stabinsky, Y. (1982) in Genetic Engineering (Setlow, J. and Hollaender, A., eds.), Vol. 4, pp. 1-17, Plenum Press, New York
- 22 Broido, M.S., Zon, G. and James, T.L. (1984) Biochem. Biophys. Res. Commun. 19, 663-670
- 23 Caruthers, M.H. (1985) Science 230, 281-285
- 24 Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 68, 499-560
- 25 Grunstein, M. and Hogness, D.S. (1975) Proc. Natl. Acad. Sci. USA 72, 3961-3965
- 26 Denhardt, D. (1966) Biochem. Biophys. Res. Commun. 23, 641-652
- 27 Sanger, F., Nicklin, S. and Coulson, A. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467
- 28 Goldberg, D.A. (1980) Proc. Natl. Acad. Sci. USA 77, 5794-5798
- 29 Wahl, G.M., Stern, M. and Starke, G.R. (1979) Proc. Natl. Acad. Sci. USA 76, 3683-3687
- 30 Borosius, J., Dull, T.J. and Noller, H.F. (1980) Proc. Natl. Acad. Sci. USA 77, 201-204
- 31 Brimacombe, R. (1981) Nature 294, 209-210
- 32 Southern, E. (1975) J. Mol. Biol. 98, 503-517
- 33 Cheng, S.M. and Schildkraut, C.L. (1980) Nucleic Acids Res. 8, 4075-4090
- 34 Nussinov, R. (1981) J. Mol. Biol. 149, 125-131
- 35 Blobel, G. and Dobberstein, B. (1975) J. Cell Biol. 67, 835-851
- 36 Nakamura, S., Iwanaga, S., Harada, T. and Niwa, M. (1976) J. Biochem. (Tokyo) 80, 1011-1021
- 37 Dayhoff, M.O., Barker, W.C. and McLaughlin, P.J. (1974) Origins Life 5, 311-330